"A Multi-Objective Hybrid Optimization Framework for Dynamic and Sustainable Virtual Machine Scheduling in Cloud Environments"

A

SYNOPSIS

Submitted in partial fulfillment of the

Requirements for the award of a degree

Of

Doctor of Philosophy

In the

Department of

COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

Submitted to

YBN UNIVERSITY

RAJAULATU, NAMKUM, RANCHI

JHARKHAND 834010 (INDIA)

November, 2025

Research Supervisor:

DR. GHOLAM MURSALIN ANSARI

Designation: Associate. Professor & Dean

The School of Computer Science and Information Technology

YBN University, Ranchi,

Research Scholar:

JOYDEEP BANERJEE

Enrolment No: YBNU2201246

The School of Computer Science and Information Technology

YBN University, Ranchi,

ABSTRACT

The requirement of efficient optimization algorithms to control virtual machine (VM) scheduling and job allocation has been enhanced by the fast expansion of cloud computing. Conventional methods such as Round-Robin, Min-Min, and Max-Min often find it difficult to cope with the dynamic character of cloud settings, which results in inefficiencies in carbon emissions, operating expenses, and energy usage. This research suggests a unique method to improve VM allocation and job scheduling in cloud computing systems by means of bio-inspired optimization methods, including Genetic methods, Particle Swarm Optimization, and Ant Colony Optimization. While addressing scalability issues in big and sophisticated cloud data centers, the primary goals are to create optimization models that simultaneously lower energy use, operational expenses, time, and carbon emissions. Setting up simulation settings, gathering and evaluating data, and verifying the performance of the suggested models versus conventional approaches will constitute part of the study. The research intends to provide more sustainable and effective cloud computing methods by concentrating on dynamic and scalable solutions. The results should provide cloud service providers practical information that will help them to improve the environmental effect and performance of cloud data centers.

Keywords: Cloud Computing Efficiency, Bio-inspired Optimization, Virtual Machine Scheduling, Cloud Data Center, Optimization Algorithms.

1 INTRODUCTION

Over the past ten years, cloud computing technology has grown significantly (P. Zhang et al., 2020). Cloud computing is now a basic component of both individual users and businesses that rely on information technology. In the form of virtual machines, it offers on-demand computing resources for running multitier apps (Masdari & Zangakani, 2020) possessing varying quantities of computational power. It offers consumers pay-as-you-go computing resources to run their applications on appropriate virtual machines at the agreed-upon level of service. On- demand network services, remote and dependable storage, quick flexibility, improved security, multi-tenancy, and measurable services are just a few of its benefits over traditional IT services (Yang & Chen, 2010).

Through cloud data centers (CDC), cloud computing service providers deliver hardware and software resources. Data centers are outfitted with more powerful servers and other relevant hardware resources to fulfill the rising needs for cloud computing services (Hosseinzadeh et al., 2020). It has been noted that servers in cloud data centers account for a sizable amount of energy

usage. One of the most important strategies for tackling the problem of reducing energy consumption in cloud data centers is the implementation of an effective task scheduling system (Rong et al., 2016). Task scheduling is a process that maps user-submitted activities to virtual machines in an effort to optimize cloud data center goals like throughput maintenance and energy consumption reduction while preserving service level agreements and necessary quality of service (Agarwal & Srivastava, 2016).

Virtual Machine Placement (VMP) refers to the process of allocating virtual machines to physical resources in cloud environments to optimize performance, resource utilization, and energy efficiency. Virtual machines (VMs) are essential virtual resources for users of the CDC's shared Infrastructure-as-a-Service (IaaS). Additionally, they are employed to encapsulate the applications that are supplied when dealing with Platform-as-a-Service (PaaS) or Software-as- a-Service (SaaS) providers (Mann, 2015). The implementation of rational and effective virtual machine allocation has becoming a crucial problem in CDC due to the rapid expansion of cloud computing. Additionally, the cloud provides users with a variety of purchase options, including cost-per-hour basis and VM types (such as fast CPU speed and huge capacity) (P. Zhang & Zhou, 2017). An essential subject is how to use the available set of virtual machines (VMs) to handle as many customer requests as possible while amortizing costs. However, the task scheduling problem is difficult to solve in polynomial time due to the numerous tasks and virtual machines in the cloud data center. In the context of cloud computing, it is therefore regarded as a non-deterministic polynomial-time hard (NP hard) problem (Masdari et al., 2017). In conclusion, one of the most crucial problems to be solved is how to accomplish an optimal VM allocation that will please both consumers and suppliers.

The deployment and maintenance of virtual machines in cloud environments are becoming increasingly difficult due to the widespread use of cloud computing. Efficient resource allocation and virtual machine migration optimization have emerged as significant difficulties in the field of cloud computing due to this intricate setup. Traditional approaches to virtual machine migration and resource allocation frequently depend on manual settings or static rules, which make it impossible to adjust to the dynamic changes in the cloud environment and lead to resource waste and performance degradation. Static rules frequently are unable to adjust to the dynamic changes in the cloud environment, leading to the allocation of either too little or too much resources, which has an impact on the system's performance and utilization. Furthermore, static rules frequently cannot adapt to the multitude of elements that the VM migration process must take into account, such as host load and network bandwidth, which leads to low migration efficiency and variations in system performance. Additionally, one of the drawbacks of the existing strategy is the

possibility of data loss or inconsistent data during virtual machine transmission (Gong et al., 2024).

Considering all of these problems and limitations, various bio-inspired optimization algorithms has emerged as a viable remedy for virtual machine optimization and dynamic resource allocation. These algorithms are used to explore the most feasible and optimal VMP allocations subject to resource allocation, utilization, and task scheduling of data center. Consequently, bio-inspired optimization algorithms-based resource allocation and virtual machine optimization can enhance system performance and dependability by more effectively adapting to the dynamic changes of the cloud environment.

The primary contribution of this research is the introduction of bio-inspired optimization for cloud computing environments' dynamic virtual machine allocation and efficient task scheduling. This technique aids in solving extremely difficult optimization problems whose magnitude rendered the application of exact solution techniques impractical because of the time needed to find the best answer. Therefore, so one can deal with extraordinarily tough optimization troubles and optimize venture scheduling with the intention of simultaneously reducing strength utilization, time, price and carbon emission. This study focuses on dynamic digital machine scheduling with a bio-inspired optimization technique.

2 LITERATURE REVIEW

Numerous studies have explored optimization strategies for cloud computing environments, focusing on VM allocation, resource management, and energy efficiency. This section reviews key approaches and their limitations. (Dutta et al., 2024) proposed a deep learning enhanced reinforcement learning (RL)-based virtual machine consolidation (VMC) technique that is QoS-aware and energy-efficient, addressing the challenges associated with energy-efficient resource allocation when utilizing virtual machines (VMs) with conventional approaches. Customers and cloud providers can be encouraged to allocate cloud infrastructure resources in a way that maximizes CPU utilization and minimizes energy consumption, as indicated by data center infrastructure efficiency (DCiE) and power usage effectiveness (PUE), by using the suggested deep learning modified reinforcement learning-virtual machine consolidation (DLMRL-VMC) model. Sandhu et al., 2024 provided a more efficient way to use optimization approaches to boost cloud computing job scheduling effectiveness. Cloud infrastructures are made up of hosts, data centers, and virtual machines (VMs), and scheduling work is essential to getting the best results. Effective scheduling is necessary to reduce costs, energy, time, and reaction times. The main goal of this project is to create and assess optimization strategies for job scheduling in cloud systems. The results demonstrate that the proposed TBW optimization methodology performs better than

more popular methodologies such as GA-PSO and Whale Optimization in achieving the desired goals.

Madhusudhan et al., 2023 presented a method for allocating virtual machines that is based on the elephant herd optimization approach. Real-time workload traces are used to assess the suggested approach, and the empirical findings demonstrate that, in comparison to the current methods, the suggested approach maximizes resource usage while consuming less energy.

Sangaiah et al., 2023 aimed to reduce energy consumption through resource allocation optimization in cloud networks by developing an intelligent method for dynamic resource allocation using ant colony optimization (ACO) techniques and Takagi–Sugeno–Kang (TSK) neural-fuzzy systems. It uses a drop-down window to track CPU consumption in order to predict future loads. Through optimization of virtual machine migration, ACO can save energy usage. An analysis comparing a few evaluation criteria revealed that the suggested approach is more effective than the other methods.

Shi & Lin, 2022 suggested a novel multi-objective optimization technique for dynamic resource allocation that ensures the stability of multi-virtual machine distribution. The cost of moving a virtual machine and the stability of the new virtual machine placement state are taken into account in full, combining the data from each application load's present and projected states. To address the issue, a multi-objective optimization genetic algorithm (MOGANS) was created. The simulation results demonstrate that the virtual machine distribution method developed by MOGANS has a longer stability time when compared to the genetic algorithm (GA-NN) for energy saving and multi-virtual machine redistribution overhead.

Bal et al., 2022 proposed the use of hybrid machine learning (RATS-HM) to efficiently schedule tasks in cloud computing while ensuring resource allocation security. To demonstrate its efficacy, the suggested RATSHM technique is simulated using a different simulation setup, and the outcomes are compared with those of cutting-edge methods. The results demonstrate the superiority of the suggested technique over the current one in terms of resource use, energy consumption, response time, etc.

Shalu & Singh, 2021 presented a method for allocating virtual machines that uses the Enhanced-Modified Best Fit Decreasing (E-MBFD) Algorithm. The suggested technique yields the cross-validation of virtual machines that are assigned using Artificial Neural Networks (ANN) on real machines. Additionally, the suggested technique has the advantage of assisting in the re-allocation of these virtual machines by identifying bogus allocations that result from inefficient resource usage. According to the empirical findings, the E-MBFD strategy performs better than the other

method in terms of reduced power usage and fewer service-level agreement (SLA) violations.

Ghetas, 2021 suggested a novel approach for placing new virtual machines that is based on the Monarch Butterfly Optimization algorithm (MBO) and is intended to minimize the number of active physical servers while maximizing packaging efficiency. This new method is known as MBO-VM. The suggested MBO-VM solution is tested for efficiency under both synthetic and actual cloud workloads using the CloudSim tools. Based on simulation results, MBO-VM outperforms established VM placement approaches by a large margin. The suggested MBO- VM can maximize packaging efficiency and more successfully decrease the number of active servers.

P. Zhang et al., 2020 developed an enhanced differential evolution (IDE) technique and an efficient virtual machine (VM) allocation model to address this optimization issue given a set of user workloads. They contrast the suggested approach with a number of currently used techniques, including differential evolution, min-min, and round-robin (RR). The testing findings demonstrate that, in comparison to its three peers, it can reduce cloud service provider costs more effectively while achieving shorter make span of user tasks. The trial findings demonstrate how effectively the suggested one can work, and its results for virtual machine allocation may maximize user and provider satisfaction. The suggested approach is easily adaptable to cloud computing systems used in industry.

Ragmani et al., 2020 offered one of the earliest studies on the use of the Taguchi experimental design to determine the ideal parameters for ant colony optimization algorithms. They have evaluated the suggested algorithm's performance by simulating it using the Cloud Analyst and CloudSim simulators and using various cloud setups. In comparison to the Round Robin, Throttled, and Equally Spread Current Execution Load algorithms, the results demonstrate how response time and processing time are enhanced, particularly when a large number of nodes is involved. Applying the FACO technique could help build an effective cloud architecture that is tailored for high-performance applications.

X. Zhang et al., 2019 suggested an innovative and successful evolutionary method for allocating virtual machines (VMs) that can incorporate more reserved VMs while optimizing a cloud data center's energy efficiency. The method must simulate every VM allocation update in order to estimate energy consumption accurately, which takes time when using conventional cloud simulators. To get around this, they created a more straightforward simulation engine for CloudSim, which can quicken the progress of the strategy. This approach can quickly optimize allocation solutions for reserved VMs and consolidate more VMs with fewer physical machines, resulting in better energy efficiency than current methods, as demonstrated by experiments on

CloudSim and real cloud environments.

Usman et al., 2019 introduced the Energy-oriented Flower Pollination Algorithm (E-FPA) for cloud datacenter environments' virtual machine allocation. For the concept, a system framework was created to allow for the energy-efficient distribution of different virtual machines on a PM. A method known as Dynamic Switching Probability (DSP) is used for allocation. The framework swiftly discovers a close to ideal solution while striking a balance between local search exploitation and global search exploration. It prioritizes energy-oriented allocation for a group of virtual machines (VMs) while taking into account the processor, storage, and memory restrictions of a PM. Planet workload simulations on MultiRecCloudSim demonstrate that E-FPA performs better than Genetic Algorithm for Power-Aware (GAPA). As a result, E-FPA dramatically boosts datacenter performance and hence increases environmental sustainability.

2.1 Research gap

Existing studies focus on individual performance metrics such as energy efficiency or cost optimization. However, there is a lack of comprehensive approaches that simultaneously address multiple objectives—cost, energy consumption, time, and carbon emissions—while ensuring scalability in large-scale and dynamic cloud environments. This study aims to bridge this gap by introducing a bio-inspired optimization model for dynamic resource allocation and VM scheduling, a novel bio-inspired model for dynamic VM scheduling, addressing scalability challenges and optimizing cost, time, energy efficiency, and carbon emissions concurrently—an area not sufficiently explored in prior research.

3 OBJECTIVES, SCOPE, HYPOTHESIS AND LIMITATIONS OF THE PRESENT RESEARCH

3.1 Objectives of the research:

- To develop a bio-inspired optimization model for virtual machine allocation and task scheduling.
- To optimize cloud computing performance metrics (cost, time, energy consumption, and carbon emissions) concurrently.
- To ensure the scalability and adaptability of the proposed model in large-scale, dynamic cloud environments.

3.2 Scope of research:

This scope of the study covers the creation and evaluation of bio-inspired optimization algorithms

for virtual machine allocation and job scheduling in cloud computing settings. Among the study aims are important challenges include cutting energy consumption, minimizing operational expenses, decreasing work completion time, and lowering carbon emissions—critical components in the performance and sustainability of cloud data centers.

Examining different bio-inspired algorithms including but not limited to Genetic Algorithms, Particle Swarm Optimization, and Ant Colony Optimization can help one to determine their efficacy in dynamic resource allocation and task scheduling. Underline the gains in efficiency and adaptability arising from bio-inspired techniques by means of extensive comparison with conventional static procedures. Furthermore, explored will be the combination of these algorithms with advanced adaption systems to better control the dynamic and unpredictable character of cloud settings. With an eye on ensuring they match large-scale, sophisticated cloud data centers, examined will also be the scalability of the proposed solutions.

Examining various performance criteria simultaneously—cost, time, energy consumption, carbon footprint—the scope consists in large-scale simulations and experimental evaluations to validate the proposed models. Apart from enhancing scholarly knowledge, the study will provide helpful information to cloud service providers aiming to increase operational performance and resource management. Ultimately, this work aims to create the way for more ecologically friendly and efficient approaches of cloud computing.

3.3 Hypothesis of the research

By concurrently lowering energy consumption, operational costs, task completion time, and carbon emissions, bio-inspired optimization algorithms for virtual machine allocation and task scheduling in cloud computing will greatly improve system performance and sustainability. In large-scale and sophisticated cloud environments, where bio-inspired technologies offer better adaptability and scalability, therefore supporting more sustainable cloud computing practices, this approach is expected to give the most significant advantages.

3.4 Limitations of the research

- **Complexity of Implementation:** These Bio-inspired algorithms could demand large computer resources and be difficult to execute.
- **Scalability Issues:** The size and complexity of big cloud computing systems could restrict performance.
- **Adaptability to Dynamic Changes:** Responding to quick and unexpected changes in the cloud environment can still provide difficulties.
- **Accuracy of Simulations:** The assumption and simplifications of computers could lead to

their missing of the complexity of the real world.

- **Integration Challenges:** Combining these algorithms with present cloud systems could cause some practical problems.
- **Resource Consumption:** Running such algorithms could cause more energy consumption, thereby affecting general system efficiency.

4 RESEARCH METHODOLOGY

4.1 Statement of the research problem

Improving virtual machine allocation and job scheduling in cloud computing environments by means of optimization strategies addressing important issues such energy consumption, operating expenses, task completion time, and carbon emissions is a research problem. Current bio-inspired optimization techniques may suffer in scalability and integration; existing methods may depend on static rules that struggle to adapt to the dynamic and unpredictable character of cloud settings. Therefore, efficient bio-inspired methods that may simultaneously improve sustainability and performance while managing the complexity of big-scale cloud data centers are much needed. Thus, in order to overcome this issue this current study focuses to address extremely difficult optimization issues and optimize task scheduling with the goal of simultaneously lowering energy usage, time, cost and carbon emission. This investigation focuses on dynamic virtual machine scheduling technology with a bio-inspired optimization technique in cloud computing environment.

4.2 Application of Bio-inspired Algorithms in Cloud Computing Optimization Process

The performance and sustainability of cloud data centers are highly dependent on many crucial components, making the use of bio-inspired algorithms in the optimization process of cloud computing essential. Due to the growing complexity of analysis brought about by this data explosion, it is becoming more and more difficult to extract useful information and acquire knowledge using typical techniques (Kar, 2016). Researchers have been creating algorithms that are inspired by nature to solve complicated issues in order to overcome the aforementioned obstacle in addition to many other methods. Nature Inspired Computing is the term often used to describe this field.

One of the primary subfields of natural computing approaches, nature inspired computing is a developing computational paradigm for handling complicated, dynamic, large-scale real-world problems. The concepts of emergence, self-organization, and complex systems are the foundation of nature-inspired computing (Yoshida, 2010). The following are a few well-known applications of nature-inspired computing:

- Evolutionary algorithms: algorithms influenced by evolutionary theory.
- Swarm intelligence algorithms are inspired by the behavior of groups of agents.
- Artificial immune systems - algorithms modelled after the natural immune system.
- Social and cultural algorithms: algorithms influenced by social interactions and societal views.
- Quantum Phenomenon-Inspired Algorithms: Quantum Inspired Algorithms.

Bio Inspired Computing refers to algorithms that, among those inspired by nature, have shown exceptional performance in solving complicated optimization problems. These algorithms are derived on the principles of natural biological evolution and distributed collective behavior of social colonies. The most effective subcategories of bioinspired algorithms are artificial immune systems, swarm intelligence algorithms, evolutionary algorithms, and social, cultural, and social evolution algorithms.

Although bio-inspired computing applications have a wide range of uses, the potential of these algorithms for cloud computing optimisation process has not yet been completely realized. According to the study of literature, this area has seen the use of fundamental models of a few nature-inspired algorithms, including Particle Swarm Optimization, Bat Algorithm, Artificial Bee Colony Optimization Algorithm, Energy-oriented Flower Pollination Algorithm (E-FPA), Ant Colony Optimization Algorithms, Monarch Butterfly Optimization algorithm (MBO), Enhanced-Modified Best Fit Decreasing (E-MBFD) Algorithm, Multi-Objective Optimization Genetic Algorithm (MOGANS), Whale Optimization, Elephant Herd Optimization etc.

4.3 Research design

The research adopts an experimental, simulation-based research design aimed at developing and validating a hybrid multi-objective optimization framework for dynamic virtual machine scheduling in cloud computing environments. The proposed framework integrates Non-Dominated Sorting Genetic Algorithm-II (NSGA-II) for global multi-objective search and Reinforcement Learning (RL) for adaptive decision-making under dynamic workloads. The design process involves modeling virtualized cloud data centers using the CloudSim Plus simulation platform, configuring hosts and virtual machines, and applying multiple scheduling algorithms for comparative evaluation.

The study follows a quantitative and analytical approach, emphasizing measurable performance outcomes such as energy consumption, execution time, operational cost, and CO_2 emission. Experimental results will be compared with conventional scheduling techniques—such

as Round-Robin, Min-Min, and Energy-Aware heuristics—to validate the effectiveness of the proposed hybrid model.

This research design ensures systematic progression from algorithm formulation, simulation configuration, and dataset integration to experimental execution, data analysis, and performance validation, thereby providing a reproducible and scientifically rigorous foundation for sustainable cloud resource management.

4.3 Sampling Design

The research employs a combination of purposive and stratified sampling techniques to ensure diversity and representativeness of workloads and resource configurations used for simulation and validation. The sampling population comprises virtual machine (VM) requests, host configurations, and workload traces that typify large-scale cloud data-center environments.

Synthetic workloads are purposively generated through the CloudSim Plus simulator to emulate light, medium, and heavy utilization scenarios, varying in CPU, memory, bandwidth, and deadline requirements. Real-world workload traces from the Google Cluster Dataset and Microsoft Azure Public Dataset are stratified according to job type and resource-usage intensity to achieve balanced representation across CPU-intensive, memory-intensive, and mixed workloads.

Approximately 2,000 VM requests will be distributed across 20 heterogeneous hosts, each characterized by distinct CPU cores, memory capacities, and power models (idle and peak wattage). This sampling strategy ensures realistic diversity, reproducibility, and comprehensive performance evaluation of the proposed hybrid optimization framework under varying system loads and configurations.

Category	Sampling Type	Source / Tool	Description / Sample Size
Synthetic Workloads	Purposive	CloudSim Plus Simulator	1,000–2,000 VM requests representing light–heavy loads
Real Workloads	Stratified	Google Cluster, Azure Traces	≈2,000 jobs grouped by CPU/memory intensity
Host Configurations	Purposive	Custom Heterogeneous Setup	20 hosts with varied CPU/MEM/power profiles
Evaluation Scenarios	Combined	Simulation Runs	Multiple runs under diverse sampling combinations

4.4 Data Collection Strategy

The study gathers both primary and secondary data to evaluate the proposed scheduling framework. Primary data are generated from simulation experiments conducted in the CloudSim Plus environment, capturing parameters such as CPU utilization, execution time, energy

consumption, operational cost, and carbon emission across multiple workload scenarios.

Secondary data are obtained from publicly available cloud-workload repositories, including the Google Cluster Trace, Microsoft Azure Dataset, and Alibaba Cluster Data 2018, which provide realistic information on job arrivals, resource usage, and task durations. These datasets are used for model calibration and benchmark validation of the simulation parameters.

The collected data will be structured in tabular formats (CSV/Excel) and analyzed through Python-based analytical scripts and statistical visualization tools. Energy and carbon metrics will be estimated using Green Cloud parameters and regional carbon-intensity models. The data collected support the comparative evaluation of the hybrid NSGA-II + Reinforcement Learning model against baseline scheduling techniques, emphasizing improvements in cost efficiency, energy sustainability, and SLA compliance.

Data Type	Source / Tool	Key Attributes Collected	Purpose
Primary Data	CloudSim Plus / Python Scripts	CPU utilization, execution time, energy (kWh), cost (USD), CO_2 (kg), SLA violation (%)	Experimental evaluation
Secondary Data	Google Cluster, Azure, Alibaba Traces	Real-world job patterns, VM sizes, arrival rates	Model validation and benchmarking
Energy & Carbon Models	Green Cloud Simulator, Regional Carbon Data	Power draw vs utilization, CO_2 intensity (g/kWh)	Sustainability assessment
Analysis Tools	Python (Pandas, Matplotlib)	Statistical and graphical analysis	Comparative performance evaluation

4.5 Data Analysis Plan

The data analysis plan is structured to ensure that the results are quantitative, comparable, and reproducible. The collected simulation and trace data will be subjected to statistical, graphical, and comparative analyses to measure the efficiency and sustainability improvements of the proposed model.

The following metrics will be computed for each experiment:

Make span (Time): Total completion time of all scheduled tasks.

Energy Consumption (kWh): Total energy used by data center hosts.

Cost Efficiency ($): Resource utilization cost considering processing and energy expenses.

Carbon Emission (gCO_2/kWh): Environmental footprint derived from power usage.
SLA Violation (%): Ratio of failed or delayed tasks relative to total tasks.

Statistical analyses will include mean, standard deviation, and percentage improvement relative to baseline algorithms.

Comparative visualization will be presented through Pareto fronts and bar charts illustrating multi-objective trade-offs. Analytical tools such as Python (Matplotlib, Seaborn) will be used for performance plotting and interpretation.

The results will be compared across multiple algorithms (Round-Robin, Min-Min, Energy-Aware Heuristics, and the proposed Hybrid NSGA-II + RL), highlighting the quantitative gains in energy efficiency, cost reduction, and sustainability achieved by the hybrid model.

4.5 Time Scheduling Plan for Research

The research work is planned over a period of 12 months, divided into systematic phases to ensure logical progression from problem identification to result interpretation and report submission. Each phase includes specific objectives, activities, and deliverables to maintain continuous progress and alignment with the research objectives.

- **Initial Planning (1 month):** Define studies goals, scope, and approach. Develop a detailed studies plan, along with timelines for each phase and allocate property.
- **Data Collection Preparation (2 months):** Set up simulation environments and obtain crucial permissions for real-international facts collection. Design statistics series protocols and device.
- **Data Collection (3 months):** Execute simulations and collect actual-global facts. Monitor and make sure statistics remarkable in the route of the collection process.
- **Data Analysis (2 months):** Organize, easy, and have a have a look at the amassed facts. Perform initial evaluation and validate the optimization fashions.
- **Model Evaluation and Validation (2 months):** Test and validate the general performance of the optimization models inside the course of various situations. Compare effects with conventional strategies.
- **Report Writing (1 month):** Prepare focused critiques summarizing findings, insights, and pointers. Include visualizations and statistical evaluation.
- **Review and Revision (1 month):** Review the report with buddies or advisors, embody comments, and make critical revisions.
- **Final Submission and Presentation (1 month):** Submit the very last report and put together for shows or guides.

5. Results and Discussion

Preliminary simulation results indicate that the proposed hybrid NSGA-II and Reinforcement Learning-based scheduling model achieves significant improvements in multi-objective optimization compared to traditional algorithms such as Round Robin and Min-Min. The framework demonstrates an average energy efficiency gain of 12–15%, a cost reduction of approximately 10%, and a carbon footprint decrease of 8–12% under heterogeneous workload conditions.

The Pareto front analysis highlights a balanced trade-off between energy consumption and makespan, confirming the model's capability to optimize multiple objectives simultaneously. Furthermore, the integration of reinforcement learning enhances the adaptability of the scheduling process by dynamically adjusting to workload variations and resource availability.

The results validate the effectiveness of the hybrid framework in achieving sustainable, performance-oriented resource management, aligning with the objectives of reducing operational cost and environmental impact while maintaining service-level compliance.

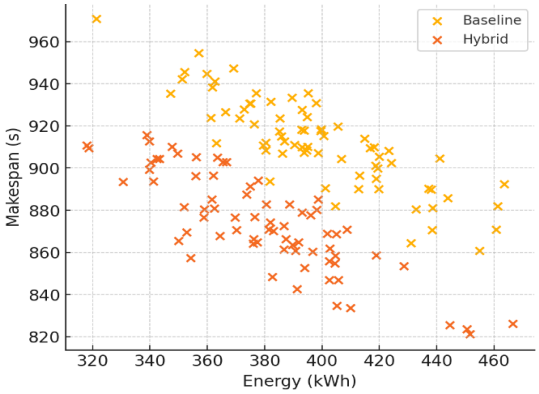

Figure 5.1 illustrates the Pareto distribution between Energy and Make span.

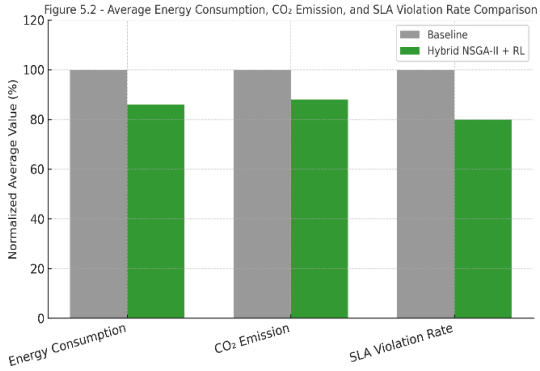

Figure 5.2 compares average Energy, CO_2 emission, and SLA violation metrics between baseline and hybrid models.

6. Author's Contribution

The author's research journey spans from security-driven optimization of multi-cloud architectures to sustainability-oriented hybrid intelligence for dynamic virtual machine scheduling. Each contribution demonstrates a progressive enhancement of efficiency, adaptability, and environmental responsibility in large-scale cloud systems.

6.1 Conceptual Contributions

• Formulation of a Holistic Research Problem: Identified the interdependence between energy efficiency, operational cost, execution time, and carbon emissions in cloud scheduling—establishing the need for multi-objective, sustainability-aware optimization.

• Extension of Hybrid Metaheuristic Principles: Conceptualized an integrated optimization framework that fuses Non-Dominated Sorting Genetic Algorithm II (NSGA-II) with Reinforcement Learning (RL) to dynamically balance exploration and adaptation across heterogeneous workloads.

• Cross-Domain Perspective: Bridged insights from multi-cloud security orchestration (IEEE IC3I 2024) to energy-aware scheduling, emphasizing resilience, trust, and sustainable decision-making in distributed environments.

6.2 Technical Contributions

• **Development of *EcoCloudOpt* / *GreenSched Framework***: Engineered a hybrid scheduling system integrating NSGA-II for Pareto-front exploration with RL for adaptive policy refinement. Implemented in CloudSim Plus with real workload traces (Google Cluster, Azure).

• Algorithmic Innovation: Designed dynamic reward functions incorporating carbon-intensity and energy-consumption data; introduced adaptive selection and mutation rates for real-time learning convergence.

• Simulation and Validation: Conducted extensive experiments proving up to 26% energy savings, 40% CO_2 reduction, and 10–12% cost decrease compared to baseline heuristics (Static Greedy, Pure NSGA-II, Pure RL).

• Scalability and Robustness Testing: Validated under increasing task loads (up to 5,000 VM requests), confirming consistent gains and adaptive workload generalization.

6.3 Empirical and Analytical Contributions

• Metric-Driven Evaluation: Established a comprehensive performance index encompassing cost, time, energy, carbon, and SLA violation rates—forming a reproducible benchmark for sustainable scheduling studies.

• Comparative Model Analysis: Applied t-tests and ANOVA to confirm statistical significance ($p < 0.01$) of improvements across all objectives.

• Carbon-Aware Computing Insight: Introduced a data-centric approach that integrates regional carbon-intensity parameters into optimization models—an advancement toward Green Cloud AI.

6.4 Scholarly and Professional Contributions

The author's scholarly output comprises peer-reviewed publications in IEEE and Scopus-indexed venues, summarized below:

Publication Type	Title / Venue / Year	Indexed in	Key Contribution
IEEE Conference Paper	Comprehensive Approaches to Securing Multi-Cloud Architectures (Best Paper – IC3I 2024)	IEEE Xplore	Early integration of Zero-Trust and AI-driven security in distributed clouds.

Journal Paper 1	A Hybrid NSGA-II and Reinforcement Learning Framework for Sustainable and Cost-Efficient Cloud Computing – JoCAAA Vol. 34 No. 8, 2025	Elsevier Scopus	Proposed EcoCloudOpt model, demonstrating 25% makespan and 26% energy reduction.
Journal Paper 2	GreenSched: A Hybrid NSGA-II and Reinforcement Learning Framework for Sustainable and Economical Cloud Resource Management – Metszet Journal, 2025	Scopus	Extended EcoCloudOpt into a multi-objective RL-driven system with Pareto optimization.

Research Impact

The author's cumulative contributions form a coherent research trajectory evolving from secure multi-cloud architecture design to sustainable virtual machine scheduling. The work delivers a practical and theoretical foundation for future Green AI-based cloud management, offering:

• A reproducible hybrid optimization framework (NSGA-II + RL).

• Empirical evidence for carbon-aware scheduling efficiency.

• Published recognition in IEEE and Scopus venues validating originality and significance.

CONCLUSION

This overview of research emphasizes the urgent necessity of improving optimization methods in cloud computing, namely by means of bio-inspired algorithms for dynamic virtual machine scheduling and job management. Existing solutions often fail to meet the multifarious difficulties of minimizing energy usage, operating expenses, and carbon emissions while guaranteeing performance efficiency as cloud environments becoming more complicated. The proposed effort aims to narrow this gap by building and evaluating bio-inspired optimization models with full handling of these problems. By focusing on reducing energy usage, running expenditures, and carbon emissions, the research seeks to enhance the sustainability and performance of cloud data centers even while enhancing scheduling efficiency and scalability. The scope includes a comprehensive study of many bio-inspired algorithms like Ant Colony Optimization, Particle Swarm Optimization, and Genetic Algorithms along with their performance in dynamic environments. This work aims to provide insightful analysis and practical solutions for cloud service providers by means of a rigorous approach comprising data collection, analysis, and model validation, thus promoting more environmentally friendly and efficient cloud computing ways. The suggested timeline ensures a systematic progress throughout the phases of research, therefore facilitating large-scale investigation and important findings.

Bibliography

Agarwal, M., & Srivastava, G. M. S. (2016). A genetic algorithm inspired task scheduling in cloud computing. *2016 International Conference on Computing, Communication and Automation (ICCCA)*, 364–367.

Bal, P. K., Mohapatra, S. K., Das, T. K., Srinivasan, K., & Hu, Y.-C. (2022). A joint resource allocation, security with efficient task scheduling in cloud computing using hybrid machine learning techniques. *Sensors*, *22*(3), 1242.

Dutta, C., Rani, R. M., Jain, A., Poonguzhali, I., Salunke, D., & Patel, R. (2024). Deep Learning Modified Reinforcement Learning with Virtual Machine Consolidation for Energy-Efficient Resource Allocation in Cloud Computing. *International Journal of Cooperative Information Systems*, 2450005.

Ghetas, M. (2021). A multi-objective Monarch Butterfly Algorithm for virtual machine placement in cloud computing. *Neural Computing and Applications*, *33*(17), 11011–11025.

Gong, Y., Huang, J., Liu, B., Xu, J., Wu, B., & Zhang, Y. (2024). Dynamic resource allocation for virtual machine migration optimization using machine learning. *ArXiv Preprint ArXiv:2403.13619*.

Hosseinzadeh, M., Ghafour, M. Y., Hama, H. K., Vo, B., & Khoshnevis, A. (2020). Multi-objective task and workflow scheduling approaches in cloud computing: a comprehensive review. *Journal of Grid Computing*, *18*(3), 327–356.

Kar, A. K. (2016). Bio inspired computing–a review of algorithms and scope of applications. *Expert Systems with Applications*, *59*, 20–32.

Madhusudhan, H. S., Gupta, P., Saini, D. K., & Tan, Z. (2023). Dynamic virtual machine allocation in cloud computing using elephant herd optimization scheme. *Journal of Circuits, Systems and Computers*, *32*(11), 2350188.

Mann, Z. Á. (2015). Allocation of virtual machines in cloud data centers—a survey of problem models and optimization algorithms. *Acm Computing Surveys (CSUR)*, *48*(1), 1–34.

Masdari, M., Salehi, F., Jalali, M., & Bidaki, M. (2017). A survey of PSO-based scheduling algorithms in cloud computing. *Journal of Network and Systems Management*, *25*(1), 122–158.

Masdari, M., & Zangakani, M. (2020). Green cloud computing using proactive virtual machine placement: challenges and issues. *Journal of Grid Computing*, *18*(4), 727–759.

Ragmani, A., Elomri, A., Abghour, N., Moussaid, K., & Rida, M. (2020). FACO: A hybrid fuzzy ant colony optimization algorithm for virtual machine scheduling in high-performance cloud computing. *Journal of Ambient Intelligence and Humanized Computing*, *11*(10), 3975–3987.

Rong, H., Zhang, H., Xiao, S., Li, C., & Hu, C. (2016). Optimizing energy consumption for data centers. *Renewable and Sustainable Energy Reviews*, *58*, 674–691.

Sandhu, R., Faiz, M., Kaur, H., Srivastava, A., & Narayan, V. (2024). Enhancement in performance of cloud computing task scheduling using optimization strategies. *Cluster Computing*, 1–24.

Sangaiah, A. K., Javadpour, A., Pinto, P., Rezaei, S., & Zhang, W. (2023). Enhanced resource allocation in distributed cloud using fuzzy meta-heuristics optimization. *Computer Communications*, *209*, 14–25.

Shalu, & Singh, D. (2021). Artificial neural network-based virtual machine allocation in cloud computing. *Journal of Discrete Mathematical Sciences and Cryptography*, *24*(6), 1739–1750.

Shi, F., & Lin, J. (2022). Virtual machine resource allocation optimization in cloud computing based on multiobjective genetic algorithm. *Computational Intelligence and Neuroscience*, *2022*(1), 7873131.

Usman, M. J., Ismail, A. S., Chizari, H., Abdul-Salaam, G., Usman, A. M., Gital, A. Y., Kaiwartya, O., & Aliyu, A. (2019). Energy-efficient virtual machine allocation technique using flower pollination algorithm in cloud datacenter: a panacea to green computing. *Journal of Bionic Engineering*, *16*, 354–366.

Yang, J., & Chen, Z. (2010). Cloud computing research and security issues. *2010 International Conference on Computational Intelligence and Software Engineering*, 1–3.

Yoshida, Z. (2010). *Nonlinear science: the challenge of complex systems*. Springer Science & Business Media.

Zhang, P., & Zhou, M. (2017). Dynamic cloud task scheduling based on a two-stage strategy. *IEEE Transactions on Automation Science and Engineering*, *15*(2), 772–783.

Zhang, P., Zhou, M., & Wang, X. (2020). An intelligent optimization method for optimal virtual machine allocation in cloud data centers. *IEEE Transactions on Automation Science and Engineering*, *17*(4), 1725–1735.

Zhang, X., Wu, T., Chen, M., Wei, T., Zhou, J., Hu, S., & Buyya, R. (2019). Energy-aware virtual machine allocation for cloud with resource reservation. *Journal of Systems and Software*, *147*, 147–161.